DOCUMENT RESUME

ED 254 551 TM 850 161

AUTHOR Thomas, R. Murray
TITLE Fiddling the Data: Vested-Interest Influence on Educational Evaluation.
PUB DATE [81]
NOTE 20p.
PUB TYPE Viewpoints (120)

EDRS PRICE MF01/PC01 Plus Postage.
DESCRIPTORS *Evaluation Methods; Evaluators; Expectation; *Integrity; Moral Values; *Program Evaluation; Program Implementation; Time Management
IDENTIFIERS *Deception; Evaluation Problems; *Evaluation Reports; Honesty; Missing Data

ABSTRACT

Evaluation errors may be unintentional, caused by incompetence, or intentional caused by mendacity. This report is concerned, not with incompetence, but with the production of educational assessments flawed by an evaluator's deceit. The purposes of the paper are: (1) to identify conditions under which people purposely distort evaluation reports, (2) to describe common techniques they use to accomplish their aims, and (3) to suggest a kind of instruction that may help prevent, or at least help uncover, the intentional falsification of educational evaluations. The observations are based on cases gathered through discussions with people engaged in educational research and development, cases appearing in the professional literature, and the author's experience in evaluation activities in the United States and in Southeast Asia and the Pacific. Circumstances which invite distorted assessments are identified, methods used to falsify assessment results are described, and a type of instruction that may encourage greater honesty is suggested. (BW)

****
* Reproductions supplied by EDRS are the best that can be made *
* from the original document. *
****

ED254551

FIDDLING THE DATA: VESTED-INTEREST INFLUENCE ON EDUCATIONAL EVALUATION

R. Murray Thomas

University of California, Santa Barbara

TM 850 161

BEST COPY

U.S. DEPARTMENT OF EDUCATION
NATIONAL INSTITUTE OF EDUCATION
EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)

☒ This document has been reproduced as received from the person or organization originating it.

☐ Minor changes have been made to improve reproduction quality.

• Points of view or opinions stated in this document do not necessarily represent official NIE position or policy.

"PERMISSION TO REPRODUCE THIS MATERIAL HAS BEEN GRANTED BY

R. M. Thomas

TO THE EDUCATIONAL RESOURCES INFORMATION CENTER (ERIC)."

FIDDLING THE DATA: VESTED-INTEREST INFLUENCE ON EDUCATIONAL EVALUATION

People with experience assessing educational projects will perhaps agree that bad evaluation, like bad government, can result when either fools or knaves are in charge. A fool's evaluation errors are unintentional, caused by incompetence. A knave's are intentional, caused by mendacity. Sometimes the two are combined, with mendacity used to cover up prior incompetence.

Of the two, incompetence is usually the easier to discover, for the evaluator is unaware of his mistakes and thus reports his procedures in a forthright manner that permits critics to identify where he has gone wrong. Mendacity is usually more difficult to expose, since the evaluator often knows the marks of sound assessment and thus is prepared to fiddle the data or to adorn the evaluation report in ways that furnish the project a facade of credibility.

Our concern in the following pages is not with incompetence but with mendacity, with the production of educational assessments flawed by an evaluator's deceit. Our three-fold purpose is (1) to identify conditions under which people --including "honest" people of "good will"— purposely distort evaluation reports, (2) to describe common techniques they use to accomplish their aims, and (3) to suggest a kind of instruction that may help prevent, or at least help uncover, the intentional falsification of educational evaluations.

The data on which the following discussion is founded have not been gathered through a controlled survey of educational projects. Indeed, the

BEST COPY

very nature of the issue makes a systematic survey impossible, since those who falsify evaluations are typicallly careful to prevent outsiders from obtaining information that would expose the deceit. It is usually only through a disgruntled or conscience-ridden insider that such information leaks out. Thus, the observations in the following pages are necessarily based on cases I have gathered casually over the past four decades through three channels —discussions with people engaged in educational research and development, cases appearing in the professional literature, and my own experience in evaluation activities both within technologically advanced cultures (principally the United States) and within developing cultures (principally in Southeast Asia and the Pacific). On the basis of these experiences, it is my impression that fiddling data is a rather widespread activity.

The comments throughout this paper are based on the assumption that the motive behind all distorted assessments is to protect or enhance someone's reputation or professional position. Much of the time this "someone" is not just the evaluator herself but also her colleagues, her bosses, an entire project staff, or an institution. In either small or large ways, the fate of many people can depend on a project's receiving a favorable evaluation, an evaluation that may require manipulated data. Thus, I propose that vested interest motivates attempts to distort evaluation reports.

A second assumption is that people —including evaluators— are not consistently honest or dishonest in all situations. Rather, honesty depends heavily on the circumstances of a given situation. Hence, I further propose that the most hopeful way to reduce the incidence of

BEST COPY

distorted assessments is not by trying to rebuild the evaluator's basic character or to quicken his conscience in general. Instead, a more practical approach is to arrange the evaluation circumstances so they discourage dishonesty. This means adjusting conditions so evaluators perceive their reputations being served better by honest reporting than by manipulating the data.

Our concern in these pages is, first, with identifying circumstances that invite distorted assessments and, second, with describing methods used to falsify assessment results. Finally, I suggest a sort of instruction that may lead to encourage greater honesty.

Circumstances That Invite Mendacity

Rather than seeking to catalogue all conditions that may affect an evaluator's honesty, I shall mention only four which often appear to influence assessments: (1) setting unrealistic expectations for an educational-development program, (2) seeking to cover up errors of planning and implementation, (3) evaluating too hastily, and (4) compensating for inadvertently lost data. These four should be enough to illustrate both a range of circumstances that occur and typical methods used for manipulating evaluation results.

Setting Unrealistic Expectations When educators or politicians predict unrealistic outcomes for an educational endeavor, they invite a future of distorted evaluations.

Various circumstances can produce such unreasonable expectations. One of the most common is that of educators competing against each other or against other agencies for the financial and political support needed to

BEST COPY

carry out their project. Under these condtions, the project's advocates are tempted to cast the expected outcomes in unduly glowing terms as they describe the venture to prospective funding agencies. Project planners are correct in assuming that the more benefits they picture as resulting from a program, the more funds financial backers are willing to provide. But what is not so apparent at the outset of a program is that judgment-day must eventually be faced when a report of the program's results comes due. On that day, when it is clear that the results have fallen short of predictions, evaluators are prone to fiddle the data and thereby close the gap between prediction and outcome.

Such overselling at the time of a project's beginning is often not an intentional act of deceit. Instead, it can result from a project leader's idealism coupled with his naivete regarding how difficult the plan will be to implement. But when the eventual evaluation is due, he recognizes that the project will still be viewed by critics as a failure, whether he pleads naivete or admits duplicity. At this point, then, the honest but naive evaluator may be willing to distort the project assessment just enough to protect the reputations of the people and institutions involved.

Unrealistic prospects can result from other circumstances as well. Sometimes an outside consultant convinces local personnel that more can be expected from a new program than actually is true. The consultant's inaccurate projections may be caused by her consciously overestimating the future outcomes in order to enhance her own image in the eyes of local educators or of funding agencies who are sponsoring her services. Or her overestimate may be unintentional, due simply to her ignorance of local conditions. Such is frequently the case when a consultant from a more

BEST COPY

developed society advises educators of a less developed society whose culture differs markedly from her own.

A further condition producing improper expectations is that of a political leader's publicly setting forth unreasonably ambitious educational outcomes in his effort to please constituents and to best his rivals. By this act he obligates educators below him in the political hierarchy to achieve the unachievable. To save their own skins, the educators are tempted to twist the evaluation results so their report of the program's success is in keeping with the leader's predictions (Thomas 1966; Thomas 1981).

In brief, a variety of situations can produce unrealistic expectations and thereby exert pressure on evaluators to distort program assessments.

Bad Planning, Bad Implementation Probably more common than unreasonable goals as stimuli for deceit are poor program planning and inefficient implementation.

One kind of poor planning involves project personnel failing to devise an adequate evaluation system at the beginning of a program. As a consequence, during the conduct of the project the data needed for a proper final evaluation are not collected. When project evaluators at last realize they lack the information they will need, they may seek to compensate for their lack of foresight either by fabricating an assessment or else by writing a report designed to mask the missing data.

Inefficient implementation means that the program is not carried through as originally envisioned. It is not that a suitable evaluation scheme was missing in the original plan but, rather, that the plan was not executed as intended. Consequently, the reputations of project personnel and perhaps

BEST COPY

of political figures are jeopardized. Again, fiddling the data may be seen as a way out of the difficulty.

Evaluating Too Hastily Sometimes pressures exerted by outside forces, such as funding agencies or community groups, urge a project staff to assess a program before it has progressed far enough to permit a fair appraisal.

For example, when inaugurating a new curriculum, staff members require sufficient time to put the curriculum into proper running order —to straighten out the initial kinks in its management— before a valid appraisal can be performed. However, if pressure groups demand reports too early, the project staff may feel compelled to protect the program by inventing an appraisal report that will stem criticism until a proper assessment can be carried out.

Conditions within granting agencies can also contribute to premature and highly positive assessments of investments in projects which are inherently of a slow-maturing variety. In effect, program officers in some granting agencies are under great internal agency pressure to protect the funding levels of their programs in competition with other officers who share the agency's overall budget. When a slow-maturing project lacks empirical indicators that demonstrate success, continued funding of the project can only be obtained by highly laudatory rhetoric. New recruits into the organization become socialized into a vocabulary of extravagant claims, and they mindlessly employ it with ever-greater hyperbole. Thus, their mendacity is not calculated distortion, but simply adaptation to an ethos and vocabulary of intensifiers, which makes them more the unintentional fool than the deliberate knave.

BEST COPY

Compensating for Lost Data Sometimes a well planned, well executed program suffers an inadvertent loss of data. Figures stored in a computer may be accidentally erased. Or a custodian may discard questionnaires that he thought were scrap paper piled in a corner. Or an epidemic causes the schools to close before the post-testing of students could be completed. Such conditions can invite project personnel to prepare an evaluation report that represents their estimate of what the data possibly would have revealed rather than telling what actually occurred.

The foregoing circumstances, then, are conditions that can lead to the publication of invalid evaluation reports. We next consider methods used for producing such a result.

Ways of Fabricating Assessments

The following are six methods that may be used for preparing biased reports.

Soldering with Rhetoric This technique consists of the evaluator seeking to repair weak data by patching the report with value judgments and data-free interpretation. For example, general laudatory or deprecatory phrases can be substituted for evidence:

"The objectives were substantially achieved in virtually all sectors of the program."

"Teachers widely agreed that pupils in the experiment made significant improvement over the school year."

Or sometimes the evaluator seeks to deflect potential criticism by prefacing the report with a rationale intended to make the subsequent assessment appear reasonable:

BEST COPY

"Instead of including complex lists of tabular results that would prove difficult for readers to analyze without the aid of a computer, we have summarized the chief outcomes in a convenient, brief descriptive form."

"Since qualititative improvement in student achievement rather than mere numerical change was the goal of the program, the following appraisal has been founded on expert judgments rather than simply on test scores that might have been compiled."

Furthermore, words intended to suggest a project's merit may be included in what ostensibly is an objective account of project outcomes. So instead of using the phrase student performance, the evaluator uses student progress or student success. Adjectives reflecting value judgments are included, as when a program's results are labeled notable, satisfactory, promising, valuable, pioneering, hoped-for, worthy, gratifying or remarkable.

Culling Out Bad Data This second method involves the evaluator's searching through the results to discover which items are favorable to the interests of the project's sponsors and which are not. The unfavorable outcomes are considered to be "bad" data and thus are culled out of the assessment. The remaining favorable results are reported. This act of selective-data-reporting may be rationalized by project personnel as necessary to serve some greater good. They feel that important ends might be jeopardized if the full data were disclosed. For instance, they may claim that monies for socially disadvantaged children, who are the objects of the program, might be curtailed and the program abandoned if negative results were included in the assessment.

Sometimes the reason data are eliminated is not to serve a greater

BEST COPY

social good nor even to support a cherished hypothesis. Instead, they are deleted to prevent embarrassment which the project staff fears it may suffer if the indeterminate or conflicting outcomes that it has found are revealed publicly. In other words, the staff fears that critics might question the project's methodology or the staff's competence if a clear-cut answer is not reported for the issue being studied.

One version of this culling approach consists of an evaluator reporting cases or anecdotal materials as if they represent the main trends in a study, whereas in reality the cases have been carefully selected to illustrate —and ostensibly to prove— the sort of result the evaluator had hoped would be found.

Featuring Desired Results Rather than culling out data, an evaluator can still bias a report by including all data in tabular or list form, then designing the accompanying narrative in a way that highlights outcomes favored by vested interests. A narrative may guide the reader's attention with such phrases as "The important findings were..." or "Of particular significance is..." or "The main point is...."

Manipulating Graphs and Tables There are several familiar ways graphs can be constructed to encourage interpretations the evaluator prefers. One is to scale the height and width of the graph in proportions that suggest a particular impression. For example, in reporting pupils' progress in foreign-language study over a period of time, the evaluator can scale language attainment up the vertical axis and time across the horizontal. Drawing the graph tall and thin gives the impression that pupils improved rapidly, whereas drawing it low and wide suggests that their progress was slow.

BEST COPY

The conclusions implied by a graph can also be affected by the draftsman's emphsizing selected trends with heavy, black lines and deemphasizing other trends by casting them in thin, light lines. In a similar way, selected figures in tables can be featured in bold-faced type.

An evaluator can further influence the impression given by a graph through selecting the occasions to be shown as key points on the graph. For example, if data have been collected on a series of occasions (once a week, once a month for two years, or each time a client appeared for a counseling session), the evaluator can have in hand more occasions than can conveniently be displayed in a table or graph. It thus may be possible for her to display only those occasions or intervals that suggest the trend she wishes to emphasize, and she leaves out those which fail to support the trend.

Choosing Self-Serving Statistics Each decade adds more varieties of stastistical treatments that can be used for describing educational data. And opinions about what are and what are not proper applications of these growing varieties abound in professional journals. An evaluator can take advantage of this expanding complex of statistical methods by selecting those which promote the conclusions he wishes to draw. If evaluators expect they may come under criticism for the statistics they have chosen, they may defend themselves by inserting in their assessment report supportive quotations from the professional literature or may mention other studies that have employed the same statistical approach.

Sometimes evaluators will influence their conclusions by choosing certain statistics, at the same time failing to include others that would be pertinent. As a simple example, reporting only the arithmetic mean and

BEST COPY

standard deviation can leave the impression that scores in a study of pupils' knowledge of science concepts were normally distributed, where in fact they were badly skewed.

In these and other ways statistics can mislead the reader from the actual outcomes of a program.

Manufacturing Data If deceit in evaluating were measured in degrees, the highest degree would likely be that of manufacturing data, that is, creating information either to fill gaps where no data exist or else to replace data that fail to suit the preferences of the evaluator or his employers. Perhaps the most celebrated example in recent years is the case of the late British psychologist, Sir Cyril Burt, accused of creating data for a fictitious nature-nurture study in his attempt to support his contention about the primacy of heredity over such environmental influences as child-raising practices and schooling (Hearnshaw 1979).

But how many other cases of manufactured data exist, on either a small or large scale, is not only unknown but likely unknowable. Of the several ways of fiddling data that we have discussed, that of fabricating false outcomes is apparently the most difficult for anyone outside of the educational project to detect. A person with some statistical background can identify many of the errors or omissions in statistical methods, graphs, and tables. And even without statistical training, a reader with some notion of logical inference can recognize when an appraisal report has been patched up with rhetoric. But a case of manufactured information is far more difficult to spot. Occasionally contraditions in the reported results alert the reader to suspect either loose interpretation or manufactured data. Or data that too perfectly fit the evaluator's original

BEST COPY

hypothesis can be a cause for suspician. However, instances of fictitious data will usually go undetected unless someone among the project's personnel is willing to expose the deception.

The foregoing six techniques, then, are ways that evaluation reports can be altered in order to lead readers to accept unwarranted conclusions that the evaluator has hoped to disseminate.

Promoting Honest Evaluating

It seems far too much to expect that the major causes behind dishonest assessment practices can be cured —causes like political ambition, a desire to protect one's job, or a wish to maintain an institution's good name. But I believe some limited progress can be made toward fostering greater honesty through a particular type of instruction. It is with such a modest expectation that the following remarks are offered.

The suggestions are directed to four sets of agents who can influence the truthfulness of assessment reports. The agents are:

1. Organizations that fund educational programs, such as legislatures, government planning bureaus, international aid bodies, and philanthropic foundations.
2. Administrators who head educational bodies and thus play an intermediary role between funding agencies and the people who actually conduct a program's activities. Examples of such administrators are ministers of education, heads of research-and-development units, and department chairmen.
3. Members of the project staff. These are the people who carry out the planning, executing, and evaluating or a program. In small projects

BEST COPY

all three functions may be performed by a single person, but more commonly the three functions are assigned to different people.

4. Professional reviewers or critics of educational programs. In some instances these people are invited by a funding agency or by an administrator to offer an outside opinion about a project. However, more often such critics have no formal relationship to a program. On their own initiative or at the invitation of a professional society or of a journal editor they take on the assignment of appraising a program, frequently publishing the results in a scholarly journal or as a review article or letter to the editor.

The role of these agents in promoting greater honesty in evaluation can be improved, I believe, if they receive instruction that focuses precisely on causes and techniques of deception and on how these can be overcome or avoided.

Instruction is recommended on the belief that the greater the technical knowledge about evaluation an agent has, the more likely that agent can play an honest role in the assessment process and not feel a need to fiddle the data. In other words, the individual will know how to perform her job without having to manipulate information dishonestly. A second reason for instructing agents is to equip them to recognize signs of deceit and thus be prepared to monitor the work of others so as to keep them honest.

Furthermore, I believe that instruction in discovering deceit can serve also as a warning, cautioning potential deceivers that other people, well prepared to dectect such dishonesty, are on the alert. Warnings are intended to inform agents of the unpleasant consequences they are likely to suffer if they twist the data. The assumption here is that people are

BEST COPY

less apt to distort evaluation reports if they believe their game may be revealed and their reputations damaged.

Instruction in Evaluation Theory and Technique

Because the four kinds of agents are expected to increase their ability for both identifying deceit and playing their own role in the evaluation process, they can profit from instruction in these two activities.

To improve their skill at recognizing deceit, they can learn about the kinds of circumstances that invite dishonesty and can study the sorts of deception techniques described earlier in this paper.

To improve their ability to play their own evaluation role honestly and to monitor the roles of other people, agents can be instructed in the stages of planning, of executing, and of evaluating educational programs, with emphasis placed on those aspects of the stages that most often affect honesty in evaluation.

It should be apparent that the kinds of instruction given may differ somewhat according to the sorts of agents receiving it. Administrators in funding organizations or those heading educational bodies (agents 1 and 2 identified above) need to know general principles rather than technical details of how to carry out the steps of the three stages. Administrators rarely have the time or patience to attend a course in research design, test construction, or curriculum planning, nor do they usually study books on these topics. However, they may be willing to attend a short workshop, participate in a conference, or read a pamphlet or an article that illustrates principles.

On the other hand, technical personnel and outside critics (agents 3 and 4) can profit from detailed knowledge of alternative ways to plan, execute,

BEST COPY

and assess. Hence, for such agents, instruction is properly in depth. What I am proposing here is not that specific courses be taught for the purpose of encouraging honesty in evaluating. Rather, I am advocating that the matters discussed in these pages be given attention within courses on evaluation, program planning, educational administration, research design, and statistics.

It seems apparent that the method of instruction most suitable on a particular occasion depends upon such conditions as the amount of time available, the learners' technical background and strength of motivation, the roles the learners as "agents" normally play, the sorts of learning materials available, and the size and location of the group being instructed. But whatever combination these conditions may form, I believe that in each case a four-step instructional procedure is appropriate. It consists of:

(1) Presenting the learners with a general problem situation, such as the issue of setting objectives for an educational project or of using rhetoric to mask shortcomings in the report of a project.

(2) Demonstrating for the learners a process for analyzing and treating the problem.

(3) Having the learners themselves practice the analysis and treatment of varied situations involving goal-setting so they achieve some active mastery of the process.

(4) Providing feedback to the learners about the adequacy of their solutions.

This four-step model can be applied in diverse ways to suit the combination of conditions found in different instructional settings. The

BEST COPY

following three examples illustrate such diversity.

Example 1. Fifty administrators and program planners in a half-day workshop engage in a sequence of five problem sessions. Each session begins with the entire group listening to an instructor explain a type of problem and illustrate one or more ways of analyzing and treating the problem. The fifty participants are then divided into groups of five members each and are furnished an array of additional problems of the same general type to analyze and treat in their group. They are also provided criteria for judging the adequacy of their solutions. One member of the group serves as an evaluator, assigned to use the criteria in assessing the quality of the group's work.

Example 2. The instructor of a university course in statistics spends one class period illustrating ways that statistics have been used to distort the true outcomes of research projects. As homework, students are provided descriptions of statistical treatments reported in a variety of projects. The students' task is to analyze the statistics and report which ones represent improper applications, why they are improper, and what should have been done to correct the condition. The students' analyses are to be written for the next class session. During that session, each student exchanges his or her homework assignment with another class member, and each writes an assessment of the other's paper, with the assessment based on a set of evaluative criteria explained during the first class session. The assessments are then passed back to the authors of the analyses, while the assignment papers themselves are handed to the instructor.

Example 3. Students in a college course in research methodology are

BEST COPY

furnished a copy of the portion of this present article that is entitled "Ways of Fabricating Assessments." They are also given segments from reports of actual or fictitious research projects. The students' assignment, as either an individual or small-group activity outside of class, is to analyze the segments in order to identify (A) what fabrication techniques they display, (B) what editing should be performed on the segments to improve their quality, and (C) what different information or additional information the authors of the reports should have provided to correct the present shortcomings of the segments. During the class session following this assignment, the instructor leads a discussion of students' analyses of each of the report segments, with class members generating criteria to use in appraising the adequacy of their analyses.

Finally we come to the matter of where instructors can obtain problem situations and case materials to serve as students' practice exercises. Whereas some cases can be found in journal articles and ERIC files, richer sources of flawed evaluation are (A) mimeographed project reports that have not gone through a refereeing process, (B) research proposals the instructor has come across while serving on a screening committee for a funding agency, (C) manuscripts the instructor has seen in his role as a reviewing agent for a journal, and (D) university students' research projects, often in the form of thesis or dissertation proposals. Or the instructor can manufacture his own fictitious cases, designing them to make their heuristic points with greatest efficiency.

REFERENCES

Hearnshaw, L.S. Cyril Burt, Psychologist. Ithaca: Cornell University

BEST COPY

Press, 1979.

Thomas, R.M. Evaluation consequences of unreasonable goals —The plight of education in American Samoa. Educational Evaluation and Policy Analysis. 1981, 3 (2), 41-49.

Thomas, R.M. Literacy by decree in Indonesia. School and Society 1966, summer, 279-283.

BEST COPY